

Making Cooperative Digital Preservation Documentation Future-Proof

Cliff Landis

Co-Chair, MetaArchive Cooperative Documentation Committee



Taming Organically Grown Documentation

Best Practices Exchange 2022

Alex Kinnaman & Kyna Herzinger
MetaArchive Technical Documentation Working Group

<https://vtechworks.lib.vt.edu/handle/10919/112181>



Cooperative Digital Preservation

MetaArchive Cooperative (2004-)



- Founded as part of NDIIPP funding and activities in 2004
- Educopia Institute founded to serve as administrative home for MetaArchive in 2006
- Currently 14 member institutions in USA, Spain, and Brazil.
- Approximately 12 TB of unique data geographically distributed across 3 continents.
- MetaArchive is a *cooperative*, not a vendor:
 - All hardware and software assets are owned by members
 - Membership fees and storage fees go to a central pool of support for network operations and members' co-op activities
- Governed *by and for members*
 - Voting representatives from all member organizations determine policies, membership costs, and network operations
- Membership Levels: Institutional: \$4,000/year ; Collaborative: \$4,000/year + \$100 per institution;
Storage fee: \$0.50/GB/year
 - Technology Rewards: If hosting a LOCKSS cache, 10% of contributed storage volume free



Cooperative of Volunteers

- Unlike some LOCKSS networks, we have no major university to centrally contribute to/manage the network.
- Salaries for community managers and technical contractors come from membership fees.
- Value: Promote a cooperative, robust, and decentralized approach to digital preservation
- Nineteen years and running!

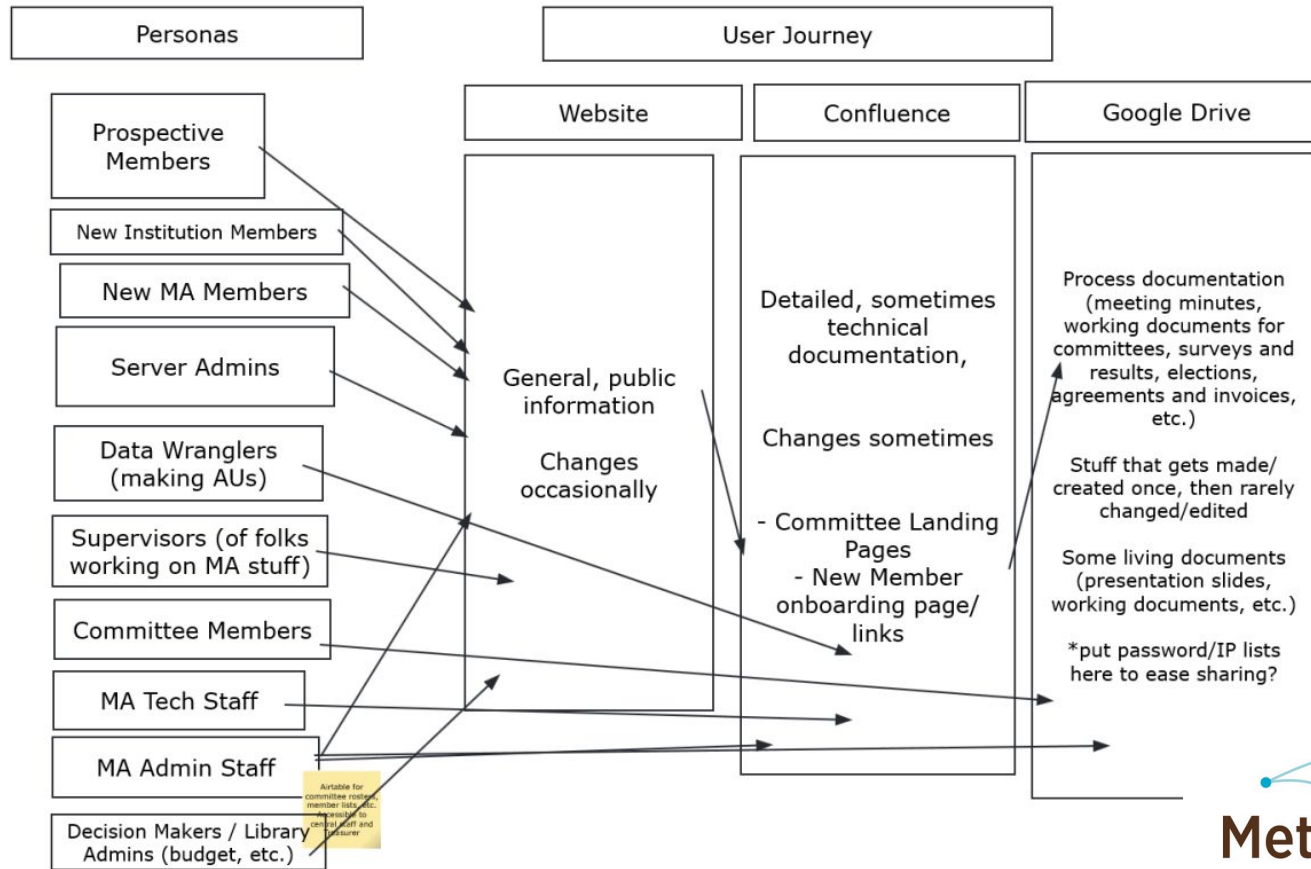
Cooperative Digital Preservation Documentation



What led up to the Documentation Working Group?

- Trust had eroded in the documentation due to several factors:
 - 15+ years of unchecked organic growth
 - Member confusion about where to find info
 - Most documentation kept behind logins
- Working Group Charge: updating documentation, and a plan for ongoing maintenance
- ~760 pages across two wikis, ~1,500 Google Drive files, website, and server-side technical documentation

Journey Mapping Visual: “Three Big Buckets”



The Low-Hanging Fruit

- Migrate meeting minutes from wikis to Google Docs
- Review Q&A Sessions in AirTable to identify immediate information needs
 - Crosslink from Confluence to AirTable and vice-versa, closing the information loop.
- Review “Orientation to MetaArchive” lesson plans to identify information needs that will impact prospective and incoming members.





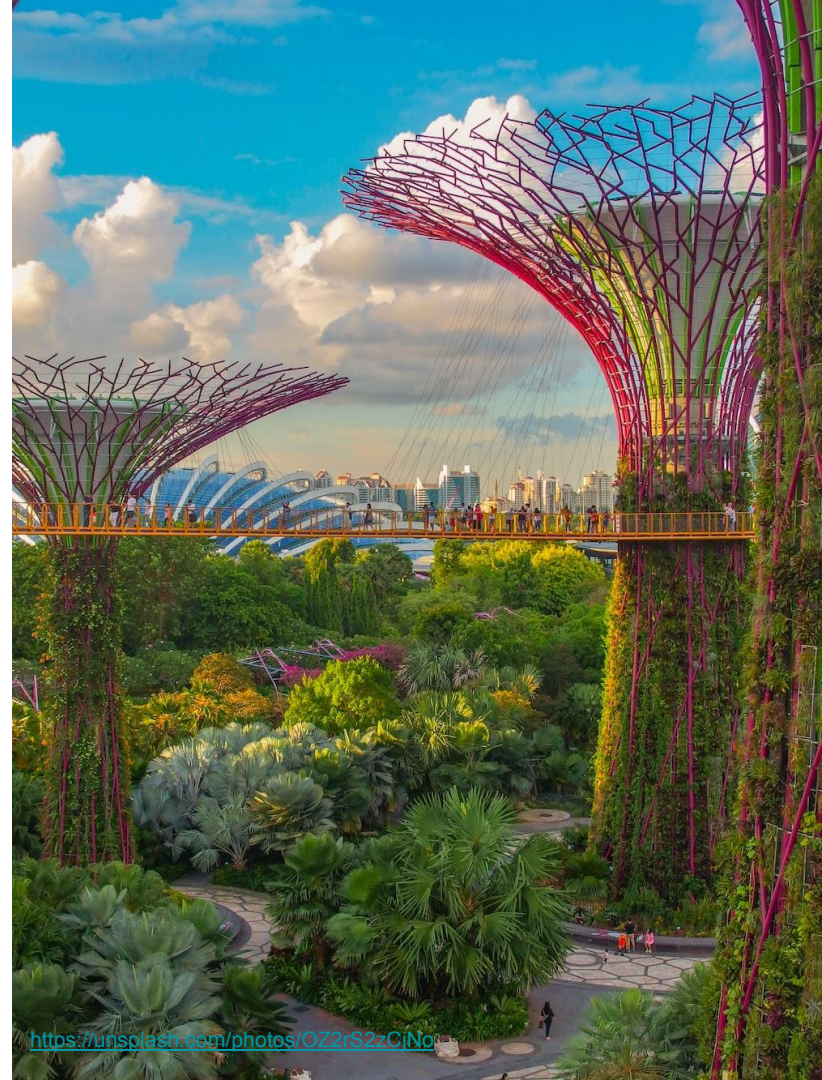
The High-Hanging Fruit: Technical Documentation

- Unlike the Website or Google Docs, the technical documentation required both technical knowledge and experience (or access to people who have it), as well as time to disentangle, update, and refine the documentation that exists.
- Blurry lines between MA tools and LOCKSS, so documenting both
- Slower work, harder to organize, and harder on morale
- Requires a willingness to be “in the thick of it”

Making it Future-Proof

How did we make it sustainable and future-proof?

- Revise the Charter to make the Documentation Working Group a permanent Committee
- Make information flow a design feature
- Pair documentation
- Make documentation as freely available as (securely) possible
- Take advantage of “working hours”

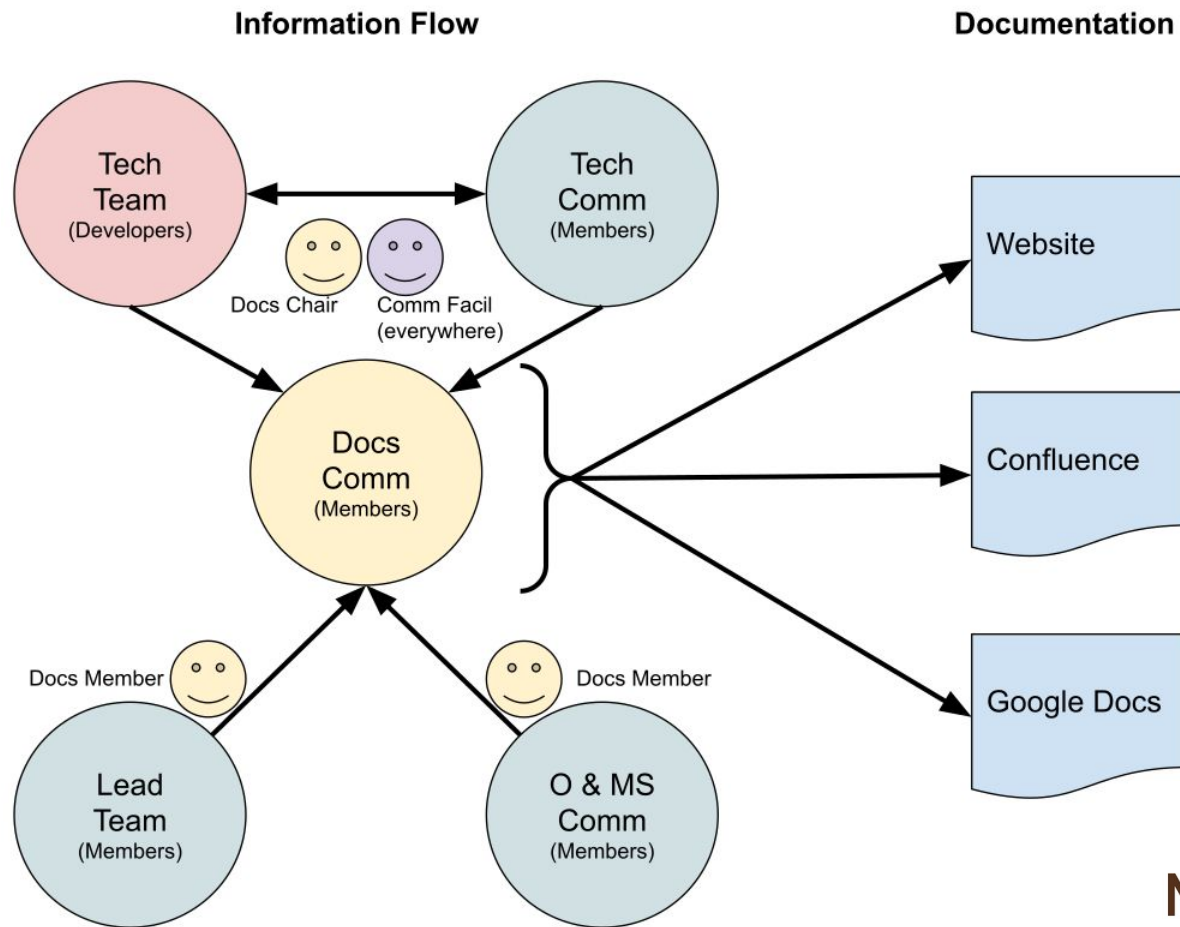


[DRAFT] Revised Charter

3.2.45. MetaArchive Documentation Committee

The Documentation Committee is responsible for developing and maintaining documentation to support the governance, outreach, and technical activities of the MetaArchive Cooperative. Documentation Committee members will liaise with MetaArchive committees and working groups to ensure that governance and outreach activities are documented; members will also liaise with the Technical Committee, Technical Team, and external technical working groups as appropriate to ensure that technical documentation is kept up-to-date.

Make Information Flow a (Committee) Design Feature



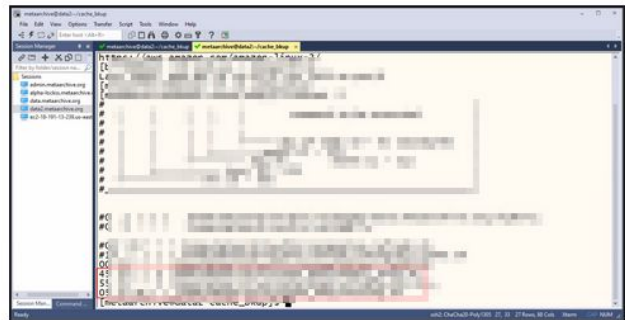
Config file backups

Created by Cliff Landis, last modified on Apr 26, 2023

- 1 Overview
- 2 Shell Scripts
 - 2.1 weekly_backup.sh
 - 2.2 validate_zip.sh
 - 2.3 mv2long.sh
- 3 Setting up a new cache's backup

Overview

This page documents the server scripts that copy the [config backups](#) from individual member's caches to the central data2 server.



- weekly_bkup.sh
- validate_zip.sh
- mv2long.sh

Shell Scripts

weekly_backup.sh

The shell script **weekly_bkup.sh** completes the weekly backup action.

Within **/home/metaarchive/cache_bkup/storage**, there is a directory for each institution's cache (illustrated here using cmu as an example):

Pair Programming Documentation

- Identify a technical documentation need, and the appropriate technical staff member
- Schedule a Zoom ride-along with technical staff
- Take screenshots of the technical staff members' screen, to capture server-based tools that I don't have access to
- Write up documentation, then ask for review by the technical staff member
- Verbal walk-through saves staff time, and allows me to write documentation for the appropriate audience

Make Documentation as Freely Available as (Securely) Possible

- Freely Available:
 - How MetaArchive Works
 - Technical Workflows
 - Knowledge Base (Glossary, Context, Best Practices)
- Keep user groups simple, based on role:
 - Most documentation publicly visible
 - Restricted areas for Technical Team (developers) and Network Administrators (IT at member institutions)
- Migrated-but-not-yet-reviewed content available to Tech Team and Network Admins





Take advantage of “working hours”

- To keep the ball rolling, we set aside dedicated co-working time, where we can work quietly together and ask questions as they come up
- It started with Documentation Working Group Working Hours on Fridays 3:00-4:00pm (rote work when you’re already brain-fried)
- The Technical Team now has 10 minutes for a quick documentation review before the main content of the meeting begins
- The Technical Committee now stays on for 30 minutes after the main meeting to work on action items from the meeting.

Observations & Future Explorations

Common Themes of Documentation & Digital Preservation

- Everyone agrees it's important, and no one wants to do it
- Treated as an afterthought
- Often the first thing to be cut when resources run out
- Requires regular maintenance work to ensure it is serving its purpose
- People turnover can bring things to a screeching halt
- Something is (almost) always better than nothing



Common Themes of Documenters & Preservationists

- Writing love letters to the future
 - “I have this frenemy called ‘Past Matt’”
- Ready to move mountains, one spoonful at a time
- Risk-averse, long-term thinkers
- View setbacks as opportunities to learn and improve workflows
- Operate in the liminal space between people and computers



“We can think of labour that goes into maintenance and repair as the work of *the maintainers*, those individuals whose work keeps ordinary existence going rather than introducing novel things. Brief reflection demonstrates that the vast majority of human labour, from laundry and trash removal to janitorial work and food preparation, is of this type: **upkeep.**”

Russell, A. & Vinsel, L. (7 April 2016) *Hail the maintainers*. Aeon.

<https://aeon.co/essays/innovation-is-overvalued-maintenance-often-matters-more>

Potential Next Steps

- Partner with LOCKSS to develop LOCKSS documentation that dovetails with MetaArchive documentation (reducing duplicative and out-of-sync documentation)
- Consolidating all code and scripts into a Git workflow (currently used for some but not all server-side coding)
- Exploring additional tools:
 - Oxford Common File Layout - <https://ocfl.io/>. A specification that describes an application-independent approach to the storage of digital information in a structured, transparent, and predictable manner. It is designed to promote long-term object management best practices within digital repositories.
 - Architecture Decision Records - <https://github.com/joelparkerhenderson/architecture-decision-record>. An architecture decision record (ADR) is a document that captures an important architecture decision made along with its context and consequences.
- Dreaming Big! Adding a dedicated Technical Writer position

Questions?

Contact MetaArchive:

Hannah Wang
MetaArchive Community Facilitator

hannah.wang@educopia.org

Connect with us:

metaarchive.org/

[@metaarchive](https://twitter.com/metaarchive)

Slides available at: <https://bit.ly/MAatBPE23>

